



PENERAPAN DATA MINING MENGGUNAKAN METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN TOKO UMAMA HIJAB KALIWATES JEMBER

Moh Sofyan Alfandi ^{a*}, Zaehol Fatah ^b

^a Sistem Informasi, sfyn.alf83@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

^b Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

*Korespondensi

ABSTRACT

Umama Hijab Kaliwates Jember Shop is a shop that specializes in selling hijabs. However, of the various hijabs being sold, not all of them sell well, some are less popular. Data on sales, purchases of goods and unexpected expenses at the Umama Hijab Store is not well structured, so the data only functions as store archives and cannot be used to develop marketing strategies. Therefore, it is necessary to apply data mining using the K-Means method at the Umama Hijab Shop. Data mining is the process of collecting and processing data to extract important information that supports decision making. The K-Means method can be applied to find out which hijab sales are selling well, selling well and not selling well. The application of this method is carried out by grouping hijab stock data. The process begins by selecting 3 random groups as initial centroids. After the data in each group did not change, the final results showed that there were 24 products that were selling well, 59 products were selling well, and 17 products were not selling well. Then, the application of the K-Means method in RapidMiner is carried out by entering product stock data, namely initial stock, sold stock and ending stock, which will be converted into a database in Ms. Excel. The data is then connected to RapidMiner Tools and processed to form groups using the K-Means algorithm. After that, RapidMiner produces product groups with high, medium and low demand.

Keywords: *Data Mining, K-Means, Umama Hijab and Rapidminer.*

Abstrak

Toko Umama Hijab Kaliwates Jember merupakan toko yang khusus menjual hijab. Namun, dari berbagai hijab yang dijual, tidak semuanya laku, ada juga yang kurang diminati. Data penjualan, pembelian barang, dan pengeluaran tak terduga di Toko Umama Hijab tidak terstruktur dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip toko dan tidak dapat digunakan untuk mengembangkan strategi pemasaran. Oleh karena itu, perlu diterapkan data mining dengan menggunakan metode *K-Means* di Toko Umama Hijab. Data mining adalah proses pengumpulan dan pengolahan data untuk mengekstrak informasi penting yang mendukung pengambilan keputusan. Metode *K-Means* dapat diterapkan untuk mengetahui penjualan hijab mana yang laku keras, laris, dan kurang laku. Penerapan metode ini dilakukan dengan cara mengelompokkan data stok hijab. Prosesnya diawali dengan memilih 3 grup acak sebagai centroid awal. Setelah data di setiap kelompok tidak berubah, hasil akhirnya menunjukkan terdapat 24 produk yang laku keras, 59 produk laris, dan 17 produk kurang laku. Kemudian, penerapan metode *K-Means* pada RapidMiner dilakukan dengan memasukkan data stok produk, yaitu stok awal, stok terjual, dan stok akhir, yang akan diubah menjadi database di Ms. Excel. Data tersebut kemudian dihubungkan ke RapidMiner Tools dan diproses untuk membentuk kelompok menggunakan algoritma *K-Means*. Setelah itu, RapidMiner menghasilkan kelompok produk yang permintaannya tinggi, sedang, dan rendah.

Kata Kunci: *Data Mining, K-Means, Umama Hijab dan Rapidminer.*

1. PENDAHULUAN

Dalam dunia bisnis yang kompetitif saat ini, masyarakat dituntut untuk selalu mengembangkan usahanya agar dapat bertahan dalam persaingan, terutama dalam persaingan penjualan, yang mendorong para pengusaha untuk menemukan pola-pola yang dapat meningkatkan penjualan dan pemasaran di dalam perusahaan. Salah satu caranya adalah melalui pemanfaatan teknologi penjualan [1]. Hal ini belum diterapkan oleh salah satu toko hijab, yaitu Toko Umama Hijab Kaliwates Jember.

Toko Umama Hijab Kaliwates Jember adalah toko yang didedikasikan untuk menjual hijab, namun tidak semua jenis hijab yang dijual memiliki tingkat penjualan yang sama. Ada produk yang sangat laris, laris, dan ada juga yang kurang laris. Namun, data penjualan, pembelian barang, dan pengeluaran tak terduga di toko ini tidak tertata dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip toko dan tidak dapat digunakan untuk pengembangan strategi pemasaran. Padahal, data yang tersedia dapat digunakan sebagai dasar dalam sistem pengambilan keputusan untuk mendukung solusi bisnis, terutama dengan memanfaatkan teknologi data mining [1].

Data mining adalah proses eksplorasi dan penggalian data untuk mendapatkan informasi yang sangat penting. Proses ini bertujuan untuk menemukan pola-pola dari data yang dapat mendukung pengambilan keputusan secara cepat dan akurat. Pola-pola tersebut dapat diperoleh dari berbagai jenis database, seperti database relasional, data warehouse, data transaksional, maupun data berorientasi objek.

Salah satu metode dalam data mining adalah K-Means, yaitu metode clustering data non-hierarkis yang mempartisi data ke dalam satu atau lebih cluster (kelompok). Data yang memiliki karakteristik serupa akan dikelompokkan ke dalam cluster yang sama [2]. Metode K-Means bekerja dengan atribut numerik dan berbasis jarak untuk membagi data menjadi beberapa cluster [3]. Dalam konteks penjualan, K-Means dapat digunakan untuk menentukan pola penjualan berdasarkan tren pembelian konsumen. Jika data dianalisis dan diolah dengan baik, metode ini dapat membantu mengetahui produk mana yang paling diminati dan produk mana yang kurang diminati. Informasi ini sangat berguna untuk mengelola inventaris toko dan sebagai masukan dalam menyusun strategi pemasaran yang lebih efektif [4].

Hasil dari data mining dapat digunakan untuk mendukung pengambilan keputusan di masa depan. “K-Means adalah metode clustering non-hierarkis yang mencoba mempartisi data menjadi satu atau lebih cluster, atau bahkan memecah data menjadi beberapa cluster [5].” Algoritma ini juga digunakan untuk memahami minat pelanggan terhadap produk yang dijual [2]. Tujuan dari penelitian ini adalah untuk membantu Toko Umama Hijab Kaliwates Jember menentukan produk hijab yang termasuk dalam kategori sangat laris, laris, dan kurang laku.

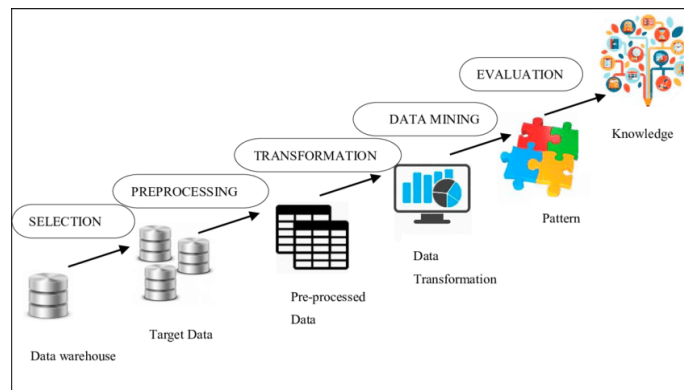
2. METODOLOGI PENELITIAN

2.1. Data Mining

Data Mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Namun tidak semua pekerjaan pencarian informasi dapat dinyatakan sebagai data mining [4]. Data mining memiliki banyak metode, metode data mining yang digunakan adalah *Knowledge Discovery in Database (KDD)* ataupun dapat disebut *Pattern Recognition*.

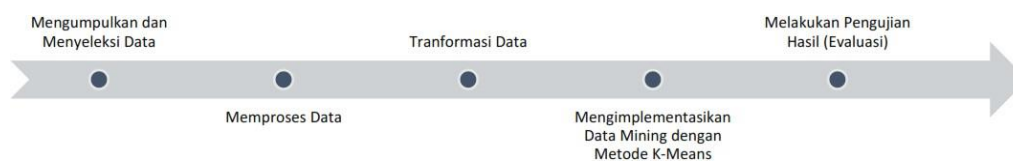
2.2. Knowledge discovery in database (KDD)

Knowledge discovery in database (KDD) adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, pola yang sudah ditemukan bersifat sah dan baru sehingga dapat bermanfaat dan dapat dimengerti. Proses dalam KDD terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap *pre-processing*, *transformasi*, *data mining* dan tahap evaluasi [5]. Tahap seleksi dilakukan untuk menargetkan data yang digunakan untuk penelitian, tahap *pre-processing* dapat dilakukan integrasi data atau penggabungan data serta dilakukan *cleaning* data, yaitu dengan menghilangkan *noise*, data *redundan*, *inkonsistensi* data, serta data yang tidak relevan, tahap transformasi adalah penggabungan data dan penyesuaian format data agar dapat diproses pada tahap data mining, tahap data mining dilakukan menggunakan algoritma yang cocok untuk permasalahan dalam data, serta tahap evaluasi yang digunakan untuk pengujian dalam data [6].



Gambar 1. Tahapan KDD

Pada gambar 1 digambarkan alur mengenai tahapan KDD, setelah disesuaikan dengan tahapan KDD dihasilkan kerangka kerja penelitian yang dijabarkan pada Gambar 2.



Gambar 2. Kerangka Kerja

2.3. Clustering Data Mining

Clustering atau klusterisasi adalah suatu teknik atau metode untuk mengelompokkan data. Menurut Tan, 2006 *clustering* adalah sebuah proses untuk pengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum.

Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain. Partisi tidak dilakukan secara manual melainkan dengan algoritma *clustering*. Oleh karena itu, *clustering* sangat berguna dan bisa menemukan *group* atau kelompok yang tidak dikenal dalam data[7].

2.4. Algoritma K-Means

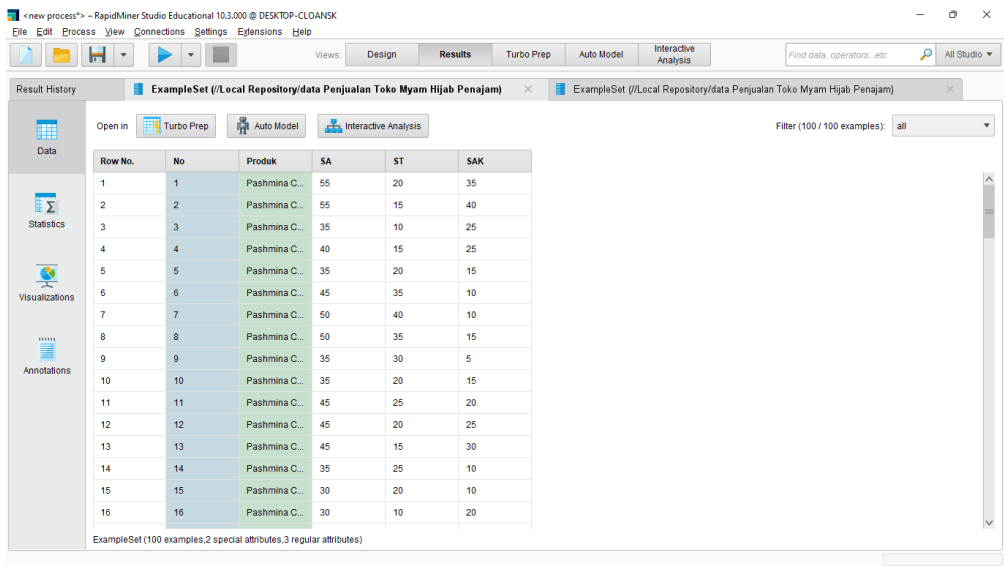
Algoritma K-Means merupakan metode yang banyak digunakan untuk mengidentifikasi suatu kelompok dari kasus. K-Means artinya prosedur pemecah pengelompokan secara iteratif yang melakukan partisi buat mengklasifikasikan atau mengelompokkan sejumlah besar objek. Prosedur pemecahan *K-Means* adalah metode *data mining* yang banyak digunakan buat mengidentifikasi suatu kelompok yang alami berasal sebuah kasus yang berdasarkan pada pengelompokan data yang memiliki kemiripan sehingga hasilnya dari pengelompokan dapat dianalisis[8]. Kelebihan dari penerapan *Algoritma K-Means* merupakan mampu mengelompokkan objek besar serta meningkatkan kecepatan proses pengelompokan

3. HASIL DAN PEMBAHASAN

Implementasi penelitian ini menggunakan perhitungan Algoritma K-Means Clustering yang dilakukan dengan tools Rapid Miner Studio. Rapid Miner sebagai tools yang mampu memudahkan user-nya dalam melakukan operator. Operator berfungsi untuk memodifikasi data dan selanjutnya dihubungkan dengan node – node operator[9].

3.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah data penjualan produk di Toko Umama Kaliwates Jember. Data ini mencakup 3 variabel utama, yaitu **SA**, **ST**, dan **SAK** yang telah ditransformasikan ke dalam format numerik agar dapat diolah lebih mudah menggunakan algoritma clustering. Dataset ini terdiri dari **100 data penjualan** yang diambil dari catatan transaksi toko. Visualisasi dataset dapat dilihat pada Gambar 3 berikut ini:

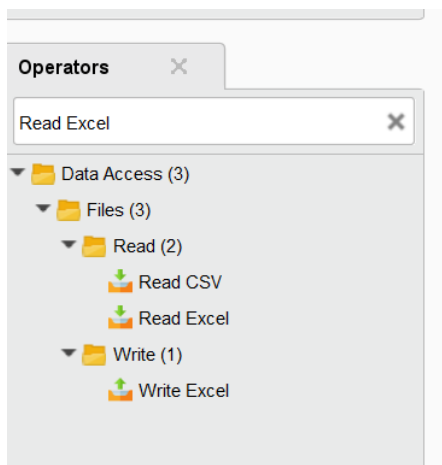


Gambar 3. Testing data- load current (amperes)

3.2 Pengolahan Data dengan software Rapidminer

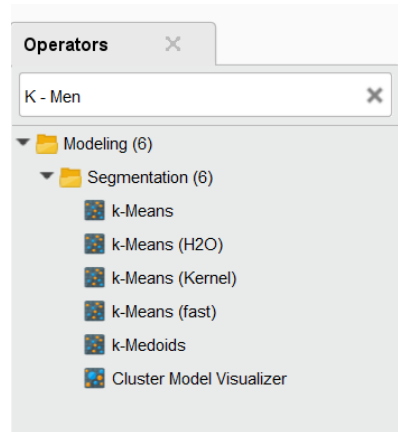
Untuk mengelompokkan data dalam penelitian ini, akan dilakukan dengan menggunakan software rapidminer 10.3 sebagai berikut:

- a. Buka *Software Rapidminer Studio Versi 10.3*.
- b. Lalu pilih menu file kemudian pilih *new proses*.
- c. Selanjutnya masuk kedalam panel operator kemudian ketik *Read Excel*, lalu drag operator *read excel* agar operator tersebut muncul di panel proses dan bisa digunakan untuk memasukkan data *excel* yang telah melalui tahap data *cleansing* dan data *transformation*. Seperti yang terlihat pada gambar 4 berikut:



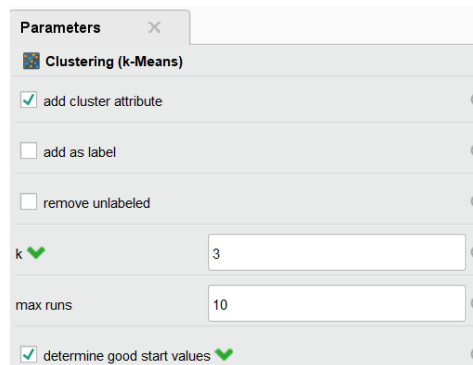
Gambar 4 Read Excel

- d. Langkah selanjutnya adalah memasukkan model yang akan digunakan untuk proses clustering data mining dengan cara masuk ke menu operator kemudian pilih K-Means seperti pada gambar 5 berikut:



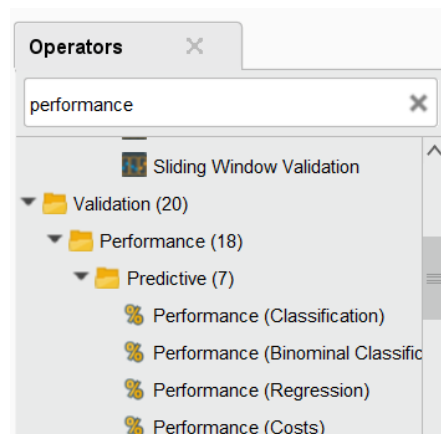
Gambar 5. Memilih Pemodelan Clustering

- e. Selanjutnya, pada proses *clustering* tentukan terlebih dahulu jumlah klasternya menjadi $K=3$ dan *mixed measure* menjadi *mixed Euclidean distance* seperti pada gambar 6 berikut:

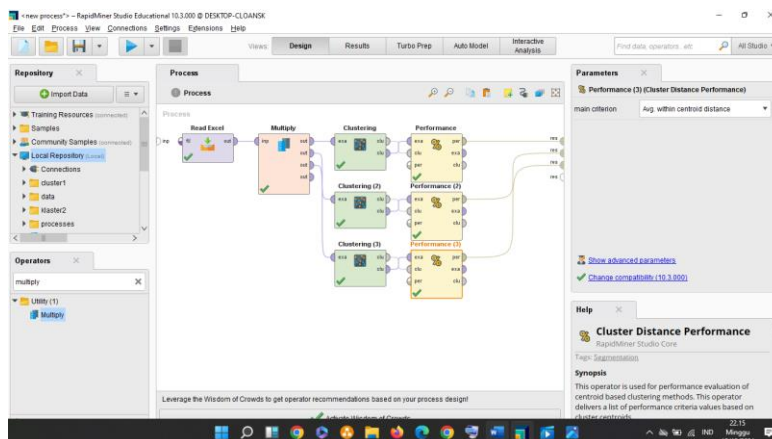


Gambar 6. Menentukan Jumlah K

- f. Langkah selanjutnya adalah memasukkan model yang akan mengukur seberapa baik hasil *clustering* yang dilakukan dengan cara masuk ke menu operator kemudian pilih *Performance* seperti pada gambar 7 berikut:

Gambar 7. Memilih Pemodelan *Performance*

- g. Selanjutnya adalah menghubungkan konektor masing – masing proses pada main process seperti pada gambar 8 dan selanjutnya klik button run untuk memulai pengklusteran.



Gambar 8. Main Proses Data Mining

3.3 Hasil Klasterisasi dengan Algoritma K-Means

a. Data View

Setelah melewati tahapan proses K-Means Clustering menggunakan aplikasi Rapidminer Studio 10.3 maka proses berikutnya adalah menampilkan hasil dari penerapan tersebut. Pada gambar 9 akan ditampilkan data view hasil clusterisasi yang telah didapatkan sebagai berikut:

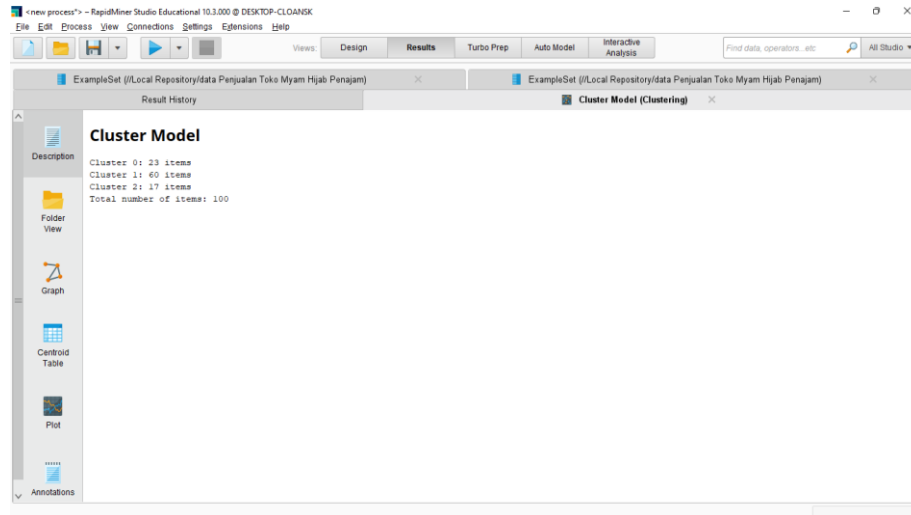
Row No.	No	Produk	cluster	SA	ST	SAK
1	1	Pashmina C.	cluster_0	55	20	35
2	2	Pashmina C.	cluster_0	55	15	40
3	3	Pashmina C.	cluster_0	35	10	25
4	4	Pashmina C.	cluster_0	40	15	25
5	5	Pashmina C.	cluster_0	35	20	15
6	6	Pashmina C.	cluster_2	45	35	10
7	7	Pashmina C.	cluster_2	50	40	10
8	8	Pashmina C.	cluster_2	50	35	15
9	9	Pashmina C.	cluster_2	35	30	5
10	10	Pashmina C.	cluster_0	35	20	15
11	11	Pashmina C.	cluster_0	45	25	20
12	12	Pashmina C.	cluster_0	45	20	25
13	13	Pashmina C.	cluster_0	45	15	30
14	14	Pashmina C.	cluster_2	35	25	10
15	15	Pashmina C.	cluster_0	30	20	10

Gambar 9. Tampilan Hasil Cluster Pada Data View

Gambar 7 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya berjumlah 100 data.

b. Cluster Model

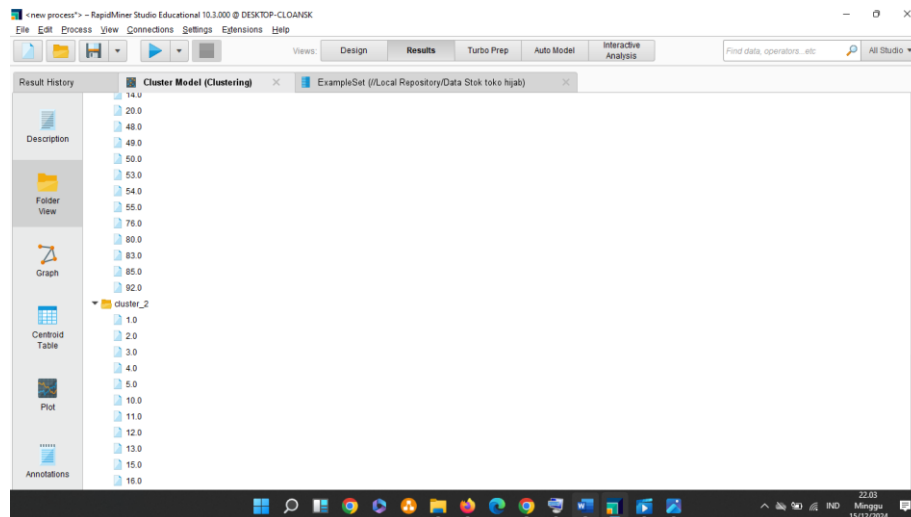
Tampilan hasil seperti gambar 10 berupa hasil pengujian dataset yang berjumlah 100 data menggunakan software rapidminer 10.3, sehingga berbentuk 3 cluster. Pada cluster 0 (cluster pertama) terdapat 23 data, cluster 1 (cluster kedua) terdapat 60 data dan cluster 2 (cluster ketiga) terdapat 17 data



Gambar 10 Cluster Model Hasil Data Mining

c. Folder View

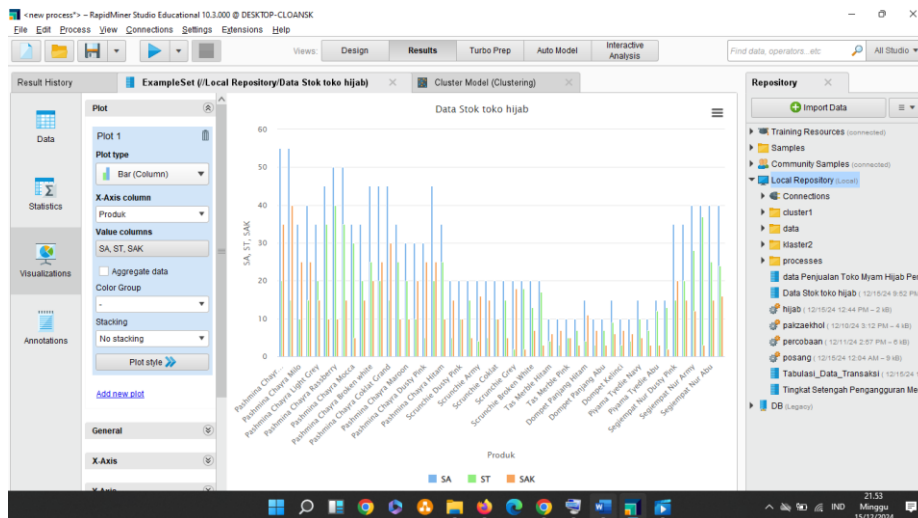
Gambar 11 di bawah ini merupakan ilustrasi tampilan Folder View yang menampilkan analisis pola penjualan di Toko Umama Kaliwates Jember. Tampilan ini memvisualisasikan data penjualan berdasarkan kategori produk dan tren penjualan selama periode tertentu, yang bertujuan untuk membantu identifikasi pola penjualan seperti produk terlaris atau waktu penjualan terbaik, sebagaimana terlihat pada Gambar 9 berikut:



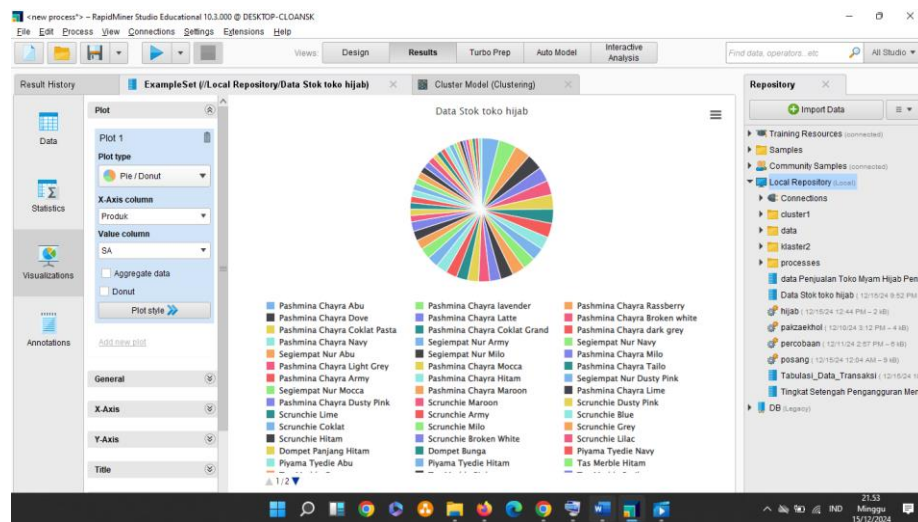
Gambar 11. Folder View

d. Visualization Hasil Clustering

Chart merupakan tampilan grafik hasil pengelompokan (clustering) data penjualan di Toko Umama Kaliwates Jember. Data dikelompokkan ke dalam 3 cluster berdasarkan analisis 6 variabel penelitian yang relevan, seperti kategori produk, frekuensi penjualan, dan margin keuntungan. Gambar 10 menampilkan hasil clustering dalam bentuk chart Bar (Column), sedangkan Gambar 12 menampilkan hasil dalam bentuk chart Pie, yang memberikan visualisasi pola penjualan secara lebih komprehensif pada gambar 10 dan 11 berikut:



Gambar 12 Tampilan Bar (column) pada Chart



Gambar 13 Tampilan Pie pada Chart

4. KESIMPULAN DAN SARAN

Sehingga dari hasil penelitian dapat disimpulkan bahwa Metode *K-Means* dapat diterapkan pada Toko Umama Hijab Kaliwates Jember untuk menentukan penjualan hijab mana yang sangat laris, laris, dan kurang laris. Penerapan metode *K-Means* pada Toko Umama Hijab dilakukan dengan cara mengelompokkan data stok hijab. Langkah pertama adalah memilih 3 cluster secara acak sebagai centroid awal. Setelah data pada setiap cluster tidak mengalami perubahan, hasil akhirnya dapat diketahui, yaitu terdapat 24 produk yang sangat laris, 59 produk yang laris, dan 17 produk yang kurang laris. Proses penerapan metode *K-Means* pada *RapidMiner* dilakukan dengan memasukkan data stok produk, yaitu stok awal, stok terjual, dan stok akhir, yang kemudian dijadikan database dalam *Ms. Excel*. Data tersebut dihubungkan ke dalam Tools *RapidMiner* untuk diolah dan dibentuk menjadi cluster menggunakan algoritma *K-Means*. Setelah itu, *RapidMiner* menghasilkan informasi tentang produk mana yang sangat laris, laris, dan kurang laris.

Pada penelitian ini masih terdapat beberapa kekurangan yang perlu diperhatikan untuk perbaikan di masa mendatang. Salah satu kekurangan adalah penelitian ini hanya menggunakan data penjualan dalam rentang waktu yang terbatas. Rentang waktu yang singkat ini dapat menyebabkan hasil penelitian kurang mewakili tren penjualan jangka panjang.

Untuk penelitian selanjutnya, disarankan untuk menggunakan data penjualan dengan periode yang lebih panjang agar hasil clustering dapat lebih akurat dan mencerminkan pola penjualan yang sebenarnya. Selain

itu, faktor-faktor lain seperti musim, promosi, atau tren mode hijab juga dapat dipertimbangkan untuk analisis yang lebih komprehensif.

Ucapan Terima Kasih

Puji dan Syukur saya panjatkan kehadirat Allah SWT. yang telah melimpahkan taufiq dan hidayahnya kepada kami, sehingga saya dapat menyelesaikan jurnal ini. Terima kasih yang sebesar-besarnya kepada dosen pembimbing yang telah mengarahkan saya dengan penuh kesabaran. Dan juga terima kasih untuk segenap dukungan, khususnya kepada teman-teman yang selalu mengingatkan saya akan terselesaikannya jurnal ini. Tanpa adanya semua dukungan ini, sulit bagi saya untuk menyelesaikan jurnal ini.

DAFTAR PUSTAKA

- [1] I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019.
- [2] Y. Yulianti, D. Y. Utami, N. Hikmah, and F. N. Hasan, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Mengetahui Minat Customer Di Toko Hijab," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 241–246, 2019.
- [3] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [4] N. Normah, B. Rifai, and P. Sari, "Algoritma Apriori Sebagai Solusi Kontrol Persediaan Suku Cadang Mobil PT. Buanasakti Aneka Motor Jakarta," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 161–168, 2020.
- [5] A. Bahar, B. Pramono, and L. H. S. Sagala, "Penentuan strategi penjualan alat-alat tattoo di studio sonyxtattoo menggunakan metode," *semanTIK*, vol. 2, no. 2, pp. 75–86, 2016.
- [6] N. Hadiza, W. N. Sari, and H. Afriyadi, "Perancangan Sistem Informasi Pengelolaan Surat Masuk dan Surat Keluar pada Kantor Dinas Pengendalian Penduduk dan Keluarga Berencana Kabupaten Sarolangun," vol. 05, no. 04, pp. 11406–11414, 2023.
- [7] E. Irwansyah and M. Faisal, *Advanced Clustering: Teori dan Aplikasi*. DeePublish, 2015. [Online]. Available: <https://books.google.co.id/books?id=8y80BgAAQBAJ>
- [8] M. Triandini, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia," *J. Inf. dan Teknol.*, vol. 3, pp. 167–173, 2021, doi: 10.37034/jidt.v3i3.120.
- [9] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- [10] Sharmila, T., & Sangeetha, R. (2021). "Optimizing Clustering Performance Using Enhanced K-Means Algorithm for Market Segmentation." *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 11(2), 15-25. doi:10.5121/ijdkp.2021.11202.